Sound Monitoring Networks New Style

Dick Botteldooren (1), Bert De Coensel (1), Damiano Oldoni (1), Timothy Van Renterghem (1) and Samuel Dauwe (2)

(1) Acoustics Group, Department of Information Technology, Ghent University, Gent, Belgium
(2) IBCN, Department of Information Technology, Ghent University, Gent, Belgium

ABSTRACT

The research project IDEA: intelligent distributed environmental assessment, is aimed at using new capabilities offered by consumer hardware to deploy sound monitoring methodologies that approach human environmental sound perception as closely as possible. The low cost noise measurement nodes deployed in the network pre-process the sound to a data set that can be transmitted continuously to the central servers. On these servers basic acoustic features are extracted and a self organizing map is trained to represent commonly co-occuring ones. Once an initial mapping is available all incoming sound is analysed and recorded when (1) the combination of features is not recognized, indicating unexpected sounds; (2) the feature set matches so closely a unit in the map that the sound can be used as a prototype. Unrecognized sounds trigger attention which leads to further training of the self organized map and potential alerts. The implementation of the above methodology resulted in a tool that allows getting detailed and useful information on the sounds that people will observe in a certain environment by mapping frequency of occurrence - and dominance of different classes of sounds that can easily be auralised for example in the context of soundscape design.

INTRODUCTION

High bandwidth communication available in most urban areas in combination with cheap and powerful computing systems such as single board computers and smart phones, allows designing a new style of noise monitoring networks. In contrast to classical noise monitoring systems, storage capacity on the measurement system itself can be limited, relying more on the communication network. In addition, the density of measurement nodes can be highly increased. This in turn allows relaxing the strict quality and reliability requirements of classical monitoring systems since cross referencing between nodes and high performance signal processing allows eliminating errors efficiently. This triggered the birth of the IDEA (Intelligent, Distributed, Environmental Assessment)-project about two years ago.

Several groups have developed the idea discussed above recently. The MESSAGE project (Bell et al., 2009) combines air pollution and noise sensors but focuses mainly on sensor network and visualisation aspects. The DREAMSys project (Barhama et al, 2009) opts for MEMs microphones as a cheap and reliable technology for deploying large sensor networks. MEMs have also been tested in IDEA but turned out to be less reliable at low temperatures (Van Renterghem et al., 2011). Also DREAMSys mainly considers hardware and network aspects. The INCAS^3 initiative in the Netherlands has a very similar scope as the network we are proposing but takes a different route in implementing it (Krijnders et al. 2010).

Designing this new style of noise monitoring requires several aspects of the network to be validated: reliability and accuracy of consumer microphones and network need to be checked but above all suitable strategies for handling the tremendous amount of data need to be designed. In this paper we briefly touch upon these aspects and focus in particular on getting more information suitable for noise control and soundscape design out of the measurements.

A DISTRIBUTED NOISE MONITORING NETWORK

Performance of cheap measurement hardware

Classical outdoor noise monitoring units cost several thousands of euros. They are built for reliability, accuracy, low noise floor, wheather resistance and low power consumption. Because they are very specific devices, they follow the fast evolution in communication and information technology less swiftly. Due to the elevated price, their use is often limited to low spatial density networks.

For urban noise monitoring, one could therefore envisage an alternative approach including large numbers of smaller and cheaper sensor nodes connected to a high performance communication backbone. Single board computers or smart phones can be bought for a few hundred euros. In urban context, where it is often possible to protect the measurement device from harsh wheather conditions and where power and internet access are usually available, they can be used for data acquisition and as a primary processing device. Consumer microphones come at a price of several euros. Since they form the core of the measurement system, they were investigated in more detail. A collection of microphones including two reference devices was placed at close proximity on the roof of a building, exposed to typical urban noise after they had been tested in lab conditions for linearity and noise floor. This field test was conducted over a period of over 6 months including winter (-7°C) and summer conditions (30°C). Several cheap microphones were identified that deviate less than 2 dBA for traffic noise. Taking into account that the difference between the two reference microphones after the same period amounted to 0.7 dBA we judged this quite acceptable (Van Renterghem et al., 2011).

Data communication reliability

The cheap single board computers (SBCs) basic functionality is to forward measurements received from the sensors to the storage servers, as illustrated in Figure 1. Although these SBCs are powerful, they are still limited in terms of storage

and networking capabilities. Therefore, several measures were taken to minimize data loss and make efficient use of the available network bandwidth.

An important aspect is the data buffering. All data coming from the attached sensors is directly stored on the onboard compact flash (CF) card. As a consequence, no data will be lost, in case of a power outage. Furthermore, several communication strategies were implemented. Firstly, it was decided to send the data in bursts instead of real-time for complexity reasons. Measurements are stored in a transmission buffer (e.g. with length of 20 sec) using files, these data caches are uploaded to the server. Only when the server sends an acknowledgement back, data is removed from the CF card. Before sending the data, a compression algorithm is used to mininimize the used bandwidth and speed up transmission.

The SBC also stores repeated 1 second wav recordings during several seconds. This allows the server to request an upload of sound recordings up to several seconds after the sound actually occurred. This latency is used by the algorithms described below to decide whether a sound is a typical example that should be stored.

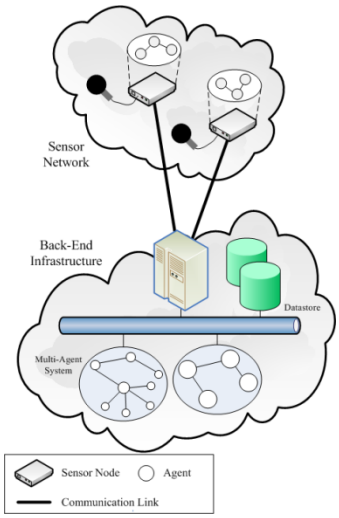

Figure 1: Overview of the network infrastructure for intelligent sound monitoring.

Agent based data analysis

Sensor networks typically fall into the category of complex, distributed, and highly dynamic systems. Because accurate and timely information is of great importance in these systems, advanced, intelligent features need to be incorporated. Multiagent systems have been recognized as one of the most suitable technologies to provide extensibility, robustness, and autonomy to the sensor network (Tynan, R., et al 2006).

In the proposed sound monitoring network, agents are used as autonomous problem solvers that co-operate to achieve a certain goal. The agents operate in a fully distributed manner, they can reside on the sensor node to exploit data locality or be located on the back-end infrastructure.

In a first stage, intelligence is added to the network at different levels to keep the measurement network operational. An important task is automatic detection of sensor or network problems. At the sensor level, cheaper and thus less accurate and reliable equipment is used. As a consequence, there is a need for identification of malfunctioning sensors. This can be achieved by monitoring each individual microphone but also by crosschecking with other sensors in the vicinity. A multi-agent system could be responsible for measurement validation (as in Amato, A et al., 2005) and malfunction identification. A positive detection can result in an alarm that will be handled by other agents.

In a second stage the human mimicking sound analysis described in the next section is implemented using agents. The tasks involved are typically very cpu intensive and become quite complex. They require more intelligent agent designs to make efficient use of the existing resources. Other agents working in parallel perform simple tasks such as updating the diurnal noise pattern or calculating hourly statistical levels. All of these agents can be activated on all or a (geographically limited) subset of the measurement nodes.

HUMAN-MIMICKING SOUND ANALYSIS

Feature extraction and saliency

The human auditory system is more sensitive to certain features of the sound than to others. Changes in time and frequency are for example more easily detected and identified. This well known fact has given rise to several feature extractors used in disciplines such as speech recognition (Cowling and Sitte, 2003). As the variety of sounds encountered in environmental noise monitoring is wider and as computational efficiency is an important aspect, more simple difference-of-Gaussian filters in time and frequency are used for extracting features. Moreover, features are computed on computation nodes and not on sensor nodes, the input data is restricted to 1/3-octave bands calculated and transmitted 8 times per second from the sensor nodes. Detection of harmonics and rhythm, two features that are prominent in human listening are not explicitly included in the feature extraction.

The features also allow estimating the saliency (Kayser et al., 2005) of the sound. Salient sounds are known to attract more human attention. While attending to daily activities or while relaxing in a tranquil environment, listening to environmental sound is seldom the primary objective. Thus, environmental sound has to attract attention before it is consciously perceived and judged either positively or negatively. Therefore systems for artificial sound analysis should account for saliency to obtain information relevant for human listening.

Mapping on the basis of temporal coherence

The feature space easily spans many dimensions. Using only 6 temporal contrast filters on 31 1/3-octave bands combined with 6 spectral contrast filters already results in a 768 dimensional feature space. To reduce dimensions, temporal coherence is used. The importance of temporal coherence in auditory scene analysis and learning in humans has recently been confirmed on a neurological basis (Shamma et al., 2011). The proposed approach is unique because it uses the local environmental sound to identify important combinations of features. As such it will become more specialised in identifying sounds that are typical for this sonic environment. Principle component analysis is insufficient to structure this multidimensional space since it only allows spanning a plane surface. Therefore, self-organizing maps (SOM) are used for extending the mapping from the multidimensional feature space to a two dimensional map (Oldoni et al., 2010). Initial training can be based on single microphone data or on a collection of data from all microphones in a certain area.

Continuous training

Mapping of the feature space based on temporal coherence during a couple of days has some drawbacks. Firstly, it was observed that a large area in the 2D mapping was reserved to what a human listener calls "silence". Listening more closely reveals different accents within this "silence" but nevertheless attributing so much of the mapping to it is neither human-like nor interesting from an environmental noise monitoring point of view. Secondly, sounds that occur only very occasionally such as fireworks or street music might be missed completely. Therefore the classical SOM algorithm was extended to include continuous training with saliency as a stimulating weight. Continuous training occurs whenever the feature vector v_i is very far from any of the nodes in the existing map: $d_{BMU}>d_{min}$, where d_{BMU} is the distance to the best matching unit in the map. Such a situation corresponds to a combination of features that has not frequently been observed during prior training. In that case a fragment of measurements is stored and used for improving the map in a parallel process. To assure more diversity in the mapping of salient sound, the learning rate α is made dependent on the saliency s_i of the sample used for training. The learning rate α is the constant that determines how fast the best matching unit moves towards the training sample i during the training process. By making α dependent on s_i, sound with low saliency will hardly induce any learning.

To illustrate the effect of continuous learning the U-matrix of the SOM before and after a continuous learning phase of several days is shown in Figure 2. The U-matrix shows the Euclidian distance between the units of the map in the original feature space. After continuous training, many of the units have migrated to distant areas in feature space to represent specific coherent feature combinations or short sounds. Much more structure is seen in the feature map. The saliency corresponding to each unit (not shown) confirms that a larger area in the map is used to represent salient sound.

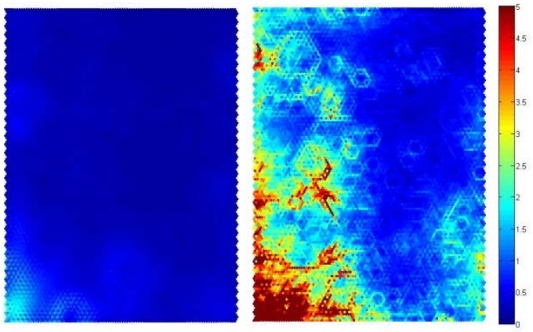

Figure 2. U-matrix of the SOM after initial training on all sound samples during 4 hours (left); after further continuous training on all samples not close to their best matching unit for 10 days.

Identifying *unfamiliar* sound

The procedure described above will automatically identify *unfamiliar* sound as sound that corresponds to a vector in feature space far away from any unit of the map. Provided that the features and the measure of saliency that were used resemble human sound perception, this will also be the sound that a human listener would recognise as unfamiliar, unusual or unexpected in the environment where the system has been trained.

If the system is allowed to train continuously, unfamiliar sound will eventually become familiar and even drift in microphone response will be accommodated. Sudden breakdown however will pop-up as an unfamiliar event.

Frequency of noticing sounds

One of the final goals of the environmental sound analysis is to discover how often specific sounds will be noticed within a given context. Attention focussing (Knudsen, 2007) plays an important role in this process. Attention can be triggered by the sound itself and in this case saliency is important or it can be outward oriented and steered by the intentions and activities of the listener. The latter is not of importance in this analysis except for the fact that some degree of continued listening after an initial triggering of attention may be desired. When implementing a model for environmental sound perception that includes attention mechanisms, accounting for inhibition of return (Spence and Driver, 1998) is essential. This mechanism prohibits that a sound with a high saliency would attract attention continuously.

Two approaches have been followed to include the above mentioned biological mechanisms in the software. The first one avoids modelling the complex attention switching mechanisms and more importantly the definition of sound objects or auditory stream. An estimate of the frequency of noticing of sound corresponding to each unit in the map (SOM) is obtained as the product $s_k.f_k$ where s_k is the saliency of unit k and f_k is the percentage of the time that unit k has been the best matching unit. Figure 3 shows an example of a map constructed at an urban measurement location where frequency of noticing is calculated for differen periods of the day.

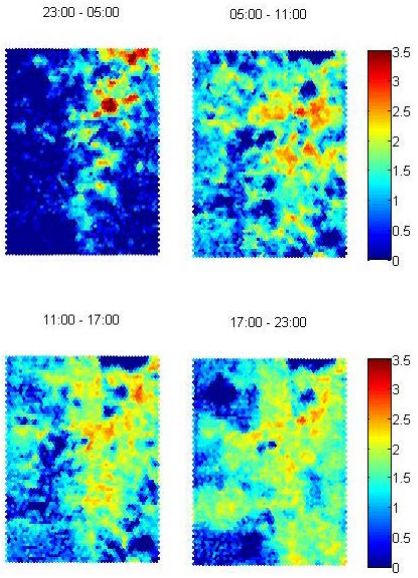

Figure 3. Estimate of frequency of noticing different sounds at different times of the day for one observation day at same location as Figure 2 on March 11, 2011.

The second and more elaborate approach models the activation and inhibition of several sounds competing for attention. For modelled sound – where the auditory streams are known by construction – this methodology has been applied in previous work (De Coensel and Botteldooren, 2010). To translate this methodology to general measured data structured in a SOM, grouping in a time retentive way of several units in the map is required. A local activation, global inhibition mechanism is required. Oscillating networks (Wang and Chang) can be used for this purpose. Although such networks gave acceptable results for the application at hand in our

previous research (Oldoni et al., 2010) they are rather CPU-time consuming. Therefore a non-oscillatory alternative is proposed. Activation of a unit k in the map is calculated as:

$$A_k(t) = f\left(d_k(t), s_k, \sum_l g(D_{l,k}, A_l(\tau)\right) \quad (1)$$

where $d_k(t)$ is the distance between the feature vector at time t and the feature vector corresponding to unit k in feature space and s_k is the saliency of unit k. $D_{l,k}$ is the distance between units l and k in the map and $A_l(\tau)$ is the activation of unit l at time τ, prior to t. The functions f and g are suitable smoothing functions. The inhibition I_k of a unit includes an inhibition of return term as well as a global inhibition term. The global inhibition assures that a limited number of units are activated at one time. As in (De Coensel and Botteldooren, 2010) also here the choice of time constants for relaxation of the activation and inhibition mechanism is crucial. Results of this extended approach are not presented in the present paper.

Labelling sounds

The very last step in the process consists in naming the different sounds. Note that the purpose here is to assign the name that the average listener would give to the sound, not the name of the sound source. This is done on the basis of an acoustic summary that selects short sound fragments that are representative for a particular area of the map. The selection procedure involves two steps. Firstly, every time step (1/8 second) the best matching unit, BMU(t), for the corresponding feature vector in the map is identified. If the distance $d_{BMU}(t)$ to this unit is shorter than any sample encountered before, the sample is identified as a prototypical example of a sound with a feature vector close to the feature vector corresponding to the particular unit. Secondly, the sound is recorded over a representative period Δt around t and stored in a sound database. A representative period Δt is defined as a period where the distance to the best matching unit at time t, $d_{BMU(t)}(t+\Delta t)$ remains within a fraction of $d_{BMU(t)}(t)$ and the distance in the map between the best matching units $D_{BMU(t),BMU(t+\Delta t)}$ is limited.

Once the sound fragments have been collected, a user can listen to the sounds using a clickable map and give them appropriate names. A panel of lay people could also be used for this, but this has not been tried so far. Figure 4 contains some of the labels associated to the sounds observed at a suburban garden. The large blue area in the bottom groups distant industrial sounds; in the upper right corner an outlier: the sound of a grass border cutted, is found; the dawn bird chorus is recovered in a central place of the map; the sound of an airplane in the distance or the sound of the windturbine is found in the bottom right area. The latter example shows the importance of context when annotating: for the average listener, this sound means airplanes, but for people living near wind turbines it is associated to the wind turbine as indeed both can hardly be distinguished from purely acoustical features. With this information – complemented with a noise level indication if needed – a representation of the local soundscape and its changes over time is obtained.

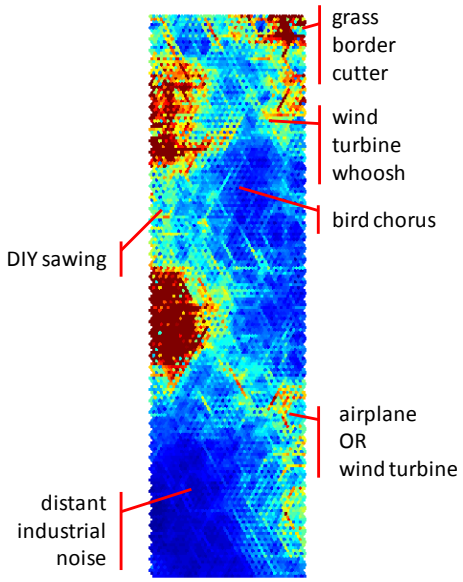

Figure 4. Annotations of some of the the sounds gathered from the system placed in a suburban garden, shown on the U-matrix; red colors indicate large distance in feature space, blue colors indicate short distance (see Figure 2).

CONCLUSIONS

A noise monitoring network based on consumer microphones and electronics backed by high performant computation servers is proposed. The backend cpu power combined with internet communication abilities and an agent based software implementation allows for new sound analysis methodologies to be deployed. A human mimicking sound classification and identification methodology was discussed.

The proposed network and sound analysis system could be useful for quantifying the importance of different sounds in the local sound climate over a longer period of time. This includes detecting rare noisy situations or even accidental sound such as explosions.

ACKNOWLEDGEMENT

The IDEA-project is a 4-year strategic basic research project, financially supported by the IWT-Vlaanderen (Flemish Agency for Innovation by Science and Technology).

REFERENCES

Amato, A. Di Lecce, V. Pasquale, C. Piuri, V. (2005), "Web agents" in an environmental monitoring system, Proc. of Computational Intelligence for Measurement Systems and Applications: 262 - 265

Barhama, R.; Goldsmith M.; Chan, M.; Simmons, D.; Trowsdale, L.; Bull, S. (2009) Development and performance of a multi-point distributed environmental noise measurement system using MEMS microphones. *Proceedings of the 8th European conference on noise control (Euronoise 2009),* Edinburgh, UK.

Bell, M.; Galatioto, F. (2009) Novel wireless pervasive sensors network to improve the understanding of noise across urban areas. *Proceedings of the 8th European conference on noise control (Euronoise 2009),* Edinburgh, UK.

Cowling M, Sitte R (2003). Comparison of techniques for environmental sound recognition. *Pattern Recognition Letters* 24(15):2895-2907.

De Coensel B, Botteldooren D (2010). A model of saliency-based auditory attention to environmental sound. *Proceedings of ICA*, Sydney, Australia.

Kayser C, Petkov C, Lippert M, Logothetis NK (2005). Mechanisms for allocating auditory attention: An auditory saliency map. *Current Biology* 15:1943-1947.

Knudsen EI (2007). Fundamental components of attention. *Annu. Rev. Neurosci.* 30:57-78.

Krijnders J.D., Niessen M.E., Andringa T.C. (2010), Sound event recognition through expectancy-based evaluation of signal-driven hypotheses, *Pattern Recognition Letters*, Volume 31, Issue 12, 1552-1559

Oldoni D, De Coensel B, Rademaker M et al. 2010. Context-dependent environmental sound monitoring using som coupled with legion. *Proc. of the IEEE International Joint Conference on Neural Networks*, Barcelona, Spain, 1413-1420.

Shamma SA, Elhilali M, Micheyl C (2011). Temporal coherence and attention in auditory scene analysis. *Trends in Neurosciences* 34(3):114-123.

Spence C, Driver J (1998). Auditory and audiovisual inhibition of return. *Percept. Psychophys.* 60(1):125-139.

Tynan, R., O'Hare, G. M. P., and Ruzzelli, A. G. (2006) Multi-agent system methodology for wireless sensor networks. Multiagent and Grid Systems, 2, 491-503.

Van Renterghem T, Thomas P., Dominguez F., Dauwe S., Touhafi A., Dhoedt B., Botteldooren D. (2011), 'On the ability of consumer electronics microphones for environmental noise monitoring', *Journal of Environmental Monitoring*, 13 (3), p. 544 - 552.

Wang D, Chang P (2008). An oscillatory correlation model of auditory streaming. *Cognitive Neurodynamics* 2(1):7-19.